

# **Modeling differences in item-position effects in the PISA 2009 reading assessment within and between schools**

Dries Debeer & Rianne Janssen (University of Leuven)  
Johannes Hartig & Janine Buchholz (DIPF)

# Content

## A. Position Effects?

1. First...
2. Two traditions
3. Combined

## B. PISA 2009 Reading assessment

1. Introduction
2. Model
3. Results
4. Discussion

# First...

## Educational measurement:

- **Common assumption:** item characteristics constant regardless ability of examinee, administration conditions
- **Common practice:** alternate test forms with different item orders
- **Common psychometric models** do not take the order / positions of items into account

# However...

Effects of **item order / item position** have been found repeatedly.

# Item Order...

## PERSON 1

Position	1	2	3	4	5	6	...	K
Item	A	B	C	D	E	F		Z

## PERSON 2

Position	1	2	3	4	5	6	...	K
Item	E	Z	F	D	B	A		C

**Different Item Orders**



# Item Position...

## PERSON 1

Position	1	2	3	4	5	6	...	K
Item	A	B	C	D	E	F		Z

## PERSON 2

Position	1	2	3	4	5	6	...	K
Item	E	Z	F	D	B	A		C

**Different Item Positions**



# However...

Effects of **item order / item position** have been found repeatedly.

## **item order:**

- Monk and Stallings (1970); MacNicol (1956); Mollenkopf (1950); Moses, Yang & Wilson (2007).

## **item position (IRT)**

- Whitely and Dawis (1976); Yen (1980); Knowles (1988); Meyers, Miller and Way (2009); Debeer and Janssen (2013)

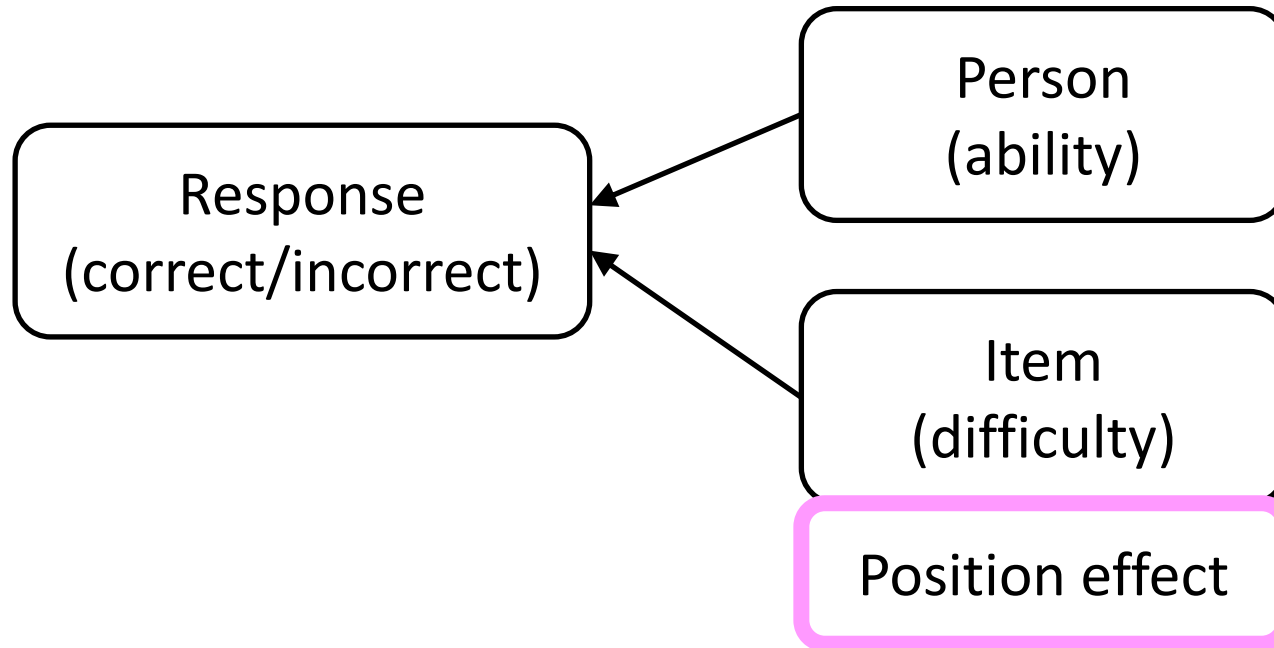
# A. Position effects?

1. Introduction
2. Methods: two traditions
  - Test side / item side
  - Person side
3. Combined approach



# Test side / item side

IRT



# Item-side

Two step approach:

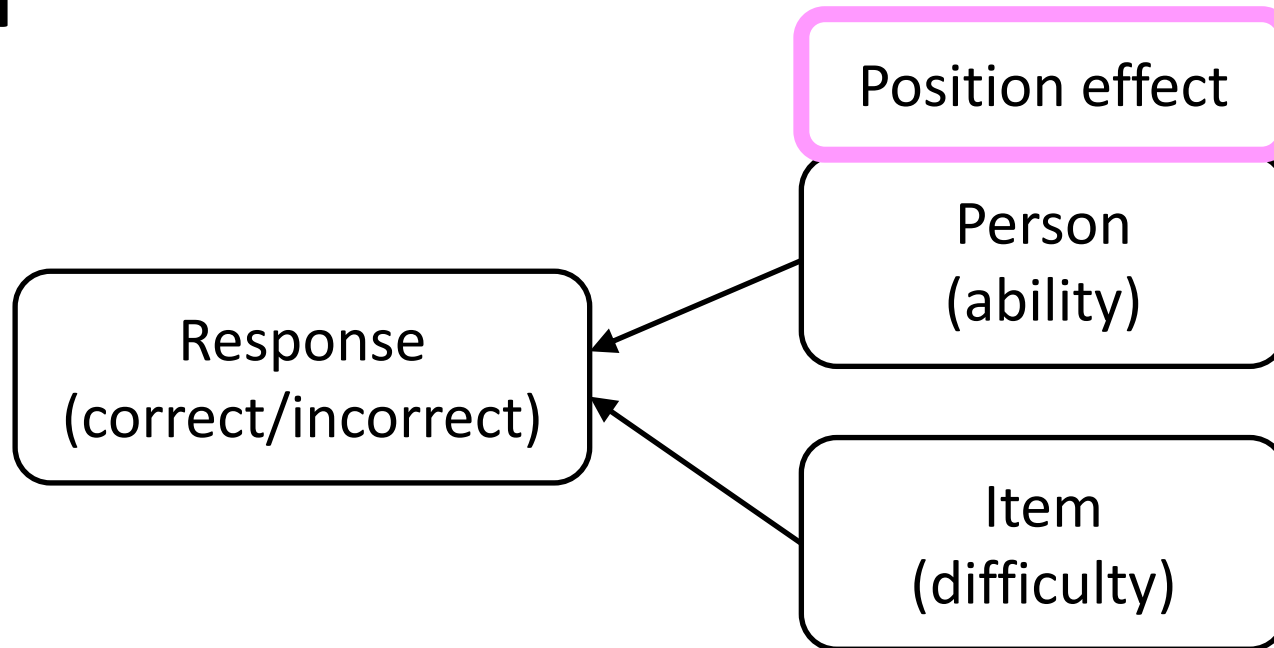
Meyers, Miller and Way (2009)

Estimate item  
difficulties within  
each test form

Relate differences  
in item difficulty  
to differences in  
item position

# Person side

IRT



# Person-side Tradition

Measurement of change

Dynamic Rasch of Verguts & De Boeck (2000)

$$\log it[P(X_{pi} = 1)] = \theta_p - t_{pi}\gamma - \beta_i$$

- $t_{pi}$  is the number of correct answers for person  $p$  up to item  $i-1$ .
- $\gamma$  is the learning parameter

# A. Position effects?

1. Introduction
2. Methods: two traditions
  - Test side / item side
  - Person side
3. Combined approach

# Combined approach

Combination of **test-side** and **person side**.  
Within an IRT-framework

# IRT-framework

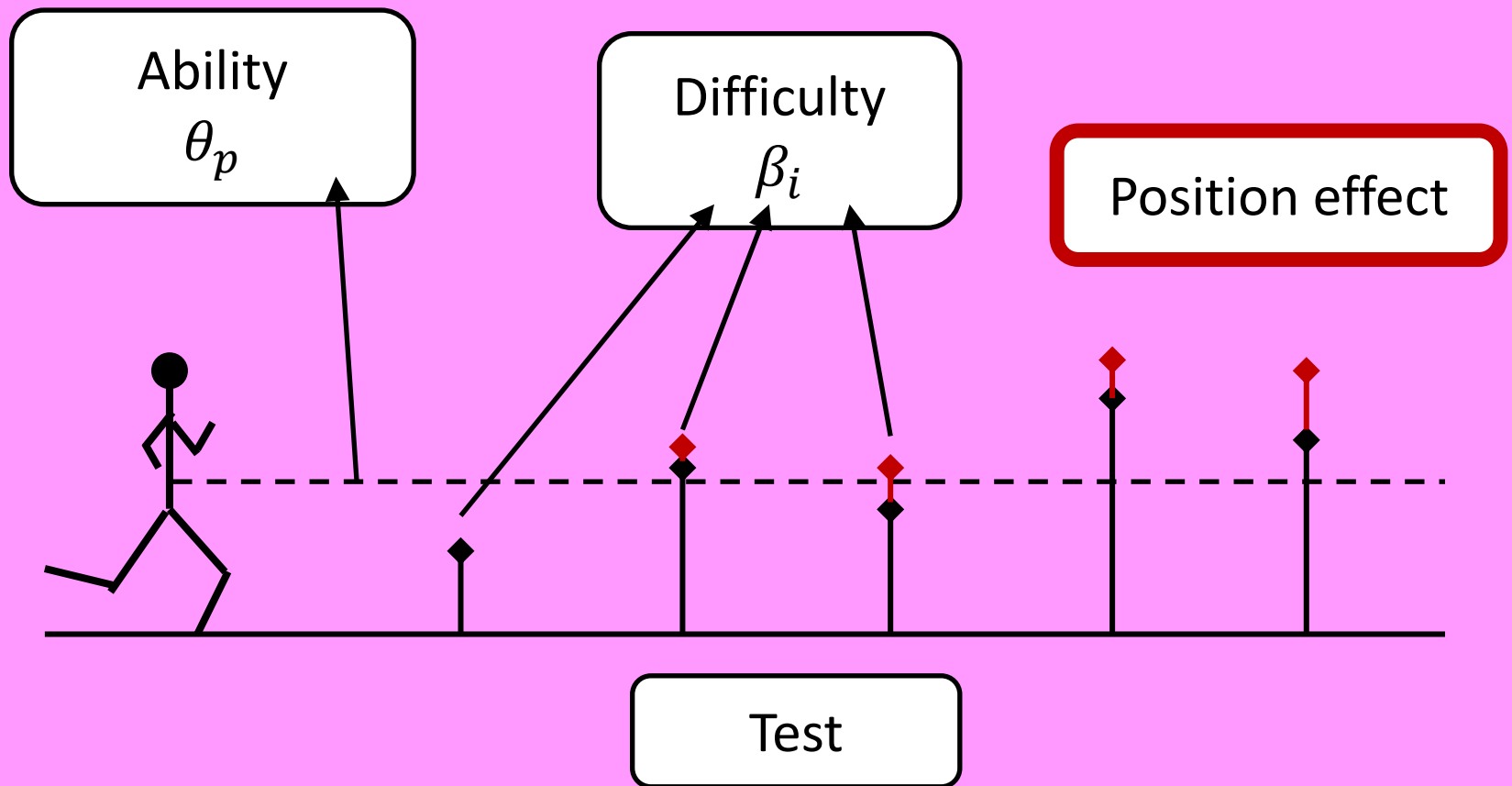
Position effect: **item side**:

$$\textit{logit} [Y_{pik} = 1] = \theta_p - \beta_i + \delta_{ik}^{\beta}$$

Position effect **across items**:

$$\textit{logit} [Y_{pik} = 1] = \theta_p - \beta_i + \delta_{\mathbf{k}}^{\beta}$$

# IRT-framework





# IRT-framework

Position effect **across items**: Trend

- **Position** as a property
- Functional form (linear effect on difficulty)

$$\text{logit} [Y_{pik} = 1] = \theta_p - \beta_i + \gamma \times (k - 1)$$

$\gamma < 0$  : a “fatigue effect”

$\gamma > 0$  : a “learning effect”

# IRT-framework

Incongruence between modeling and interpretation

⇒ Modeled at item side, interpretation at person side.

⇒ **Person side:** individual differences/position dimension

⇒ “persistence?”

# IRT-framework

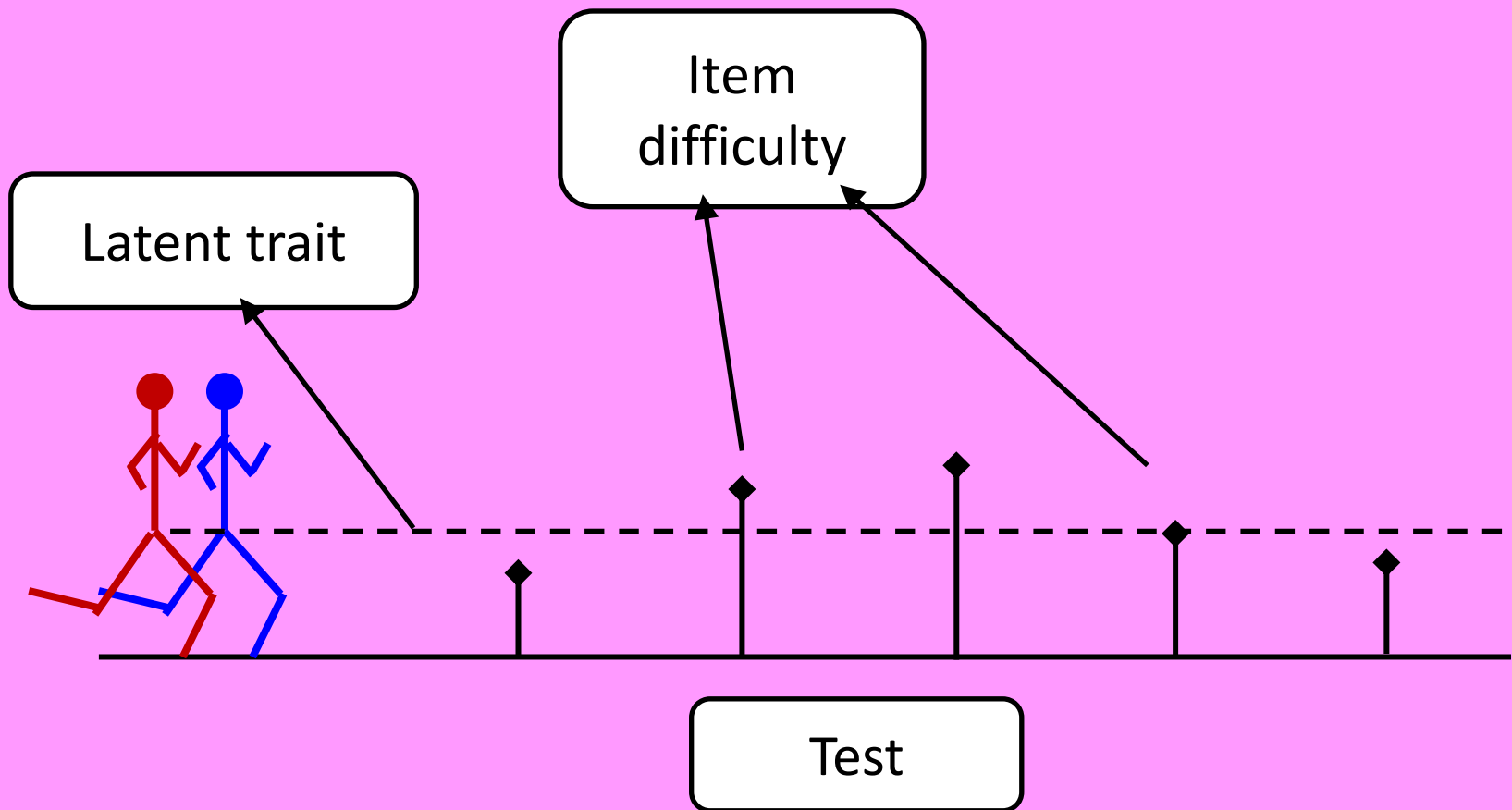
Individual differences (Person side)

$$\textit{logit} [Y_{pik} = 1] = \theta_p - \beta_i + \gamma_{\textcolor{red}{p}} \times (k - 1)$$

Interpretation!

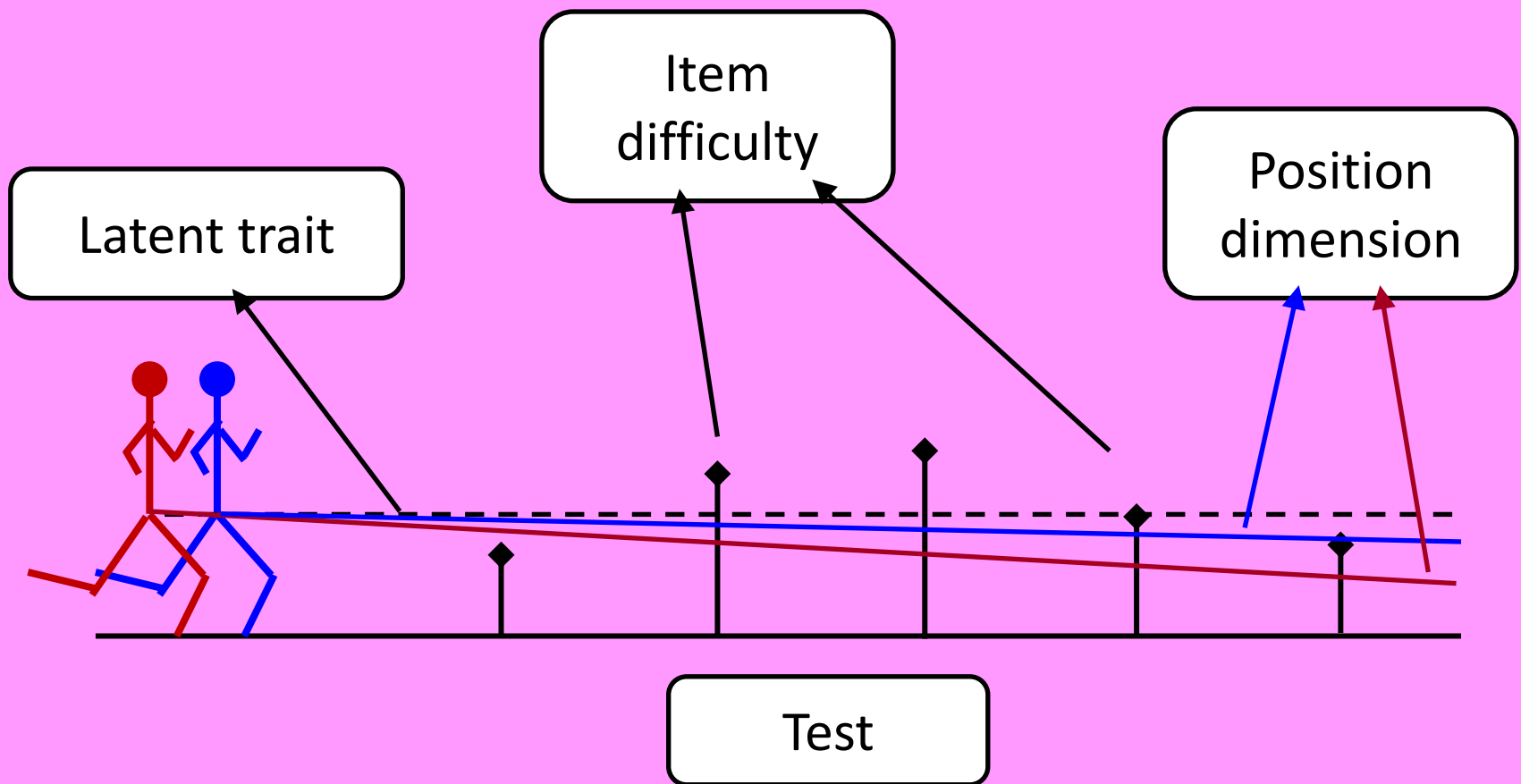
# Individual differences

## Position dimension “persistence ?”



# Individual differences

## Position dimension “persistence ?”



# Content

## A. Modeling

1. Introduction
2. Two traditions
3. Combined

## B. PISA 2009 Reading assessment

1. Introduction
2. Model
3. Results
4. Discussion

## **B. PISA 2009 Reading assessment**

- Analysis of the data form international PISA 2009 study
- In collaboration with Johannes Hartig & Janine Buchholz (DIPF)

# **B. PISA 2009 Reading assessment**

## **1. Introduction**

- PISA
- Multilevel decomposition
- Research questions

## **2. Model**

## **3. Results**

## **4. Discussion**



# Introduction

## PISA

- Program for International Student Assessment of the Organization for Economic Co-operation and Development (OECD)
- International assessment among 15-year-olds
  - Reading (R)
  - Math (M)
  - Science (S)
- Rotated cluster design
  - => Analysis of item position effect (on difficulty)

# PISA data

- Published reading assessment data from the PISA 2009 study (OECD, 2010).
- Data from 467.819 students from 16.850 schools in 65 countries was analysed.
- Responses to 125 reading test items were used in the analysis, responses on 7 partial credit items were dichotomized.

# Booklet Design

## Rotated cluster design PISA 2009

Booklet	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	M1	R1	R3	M3
2	R1	S1	R4	R7
3	S1	R3	M2	S3
4	R3	R4	S2	R2
5	R4	M2	R5	M1
6	R5	R6	R7	R3
7	R6	M3	S3	R4
8	R2	M1	S1	R6
9	M2	S2	R6	R1
10	S2	R5	M3	S1
11	M3	R7	R2	M2
12	R7	S3	M1	S5
13	S3	R2	R1	R5

# PISA 2009 Booklet Design

- The cluster (coded 0 to 3) position was used as proxy for the item position in the analyses.

## Booklet

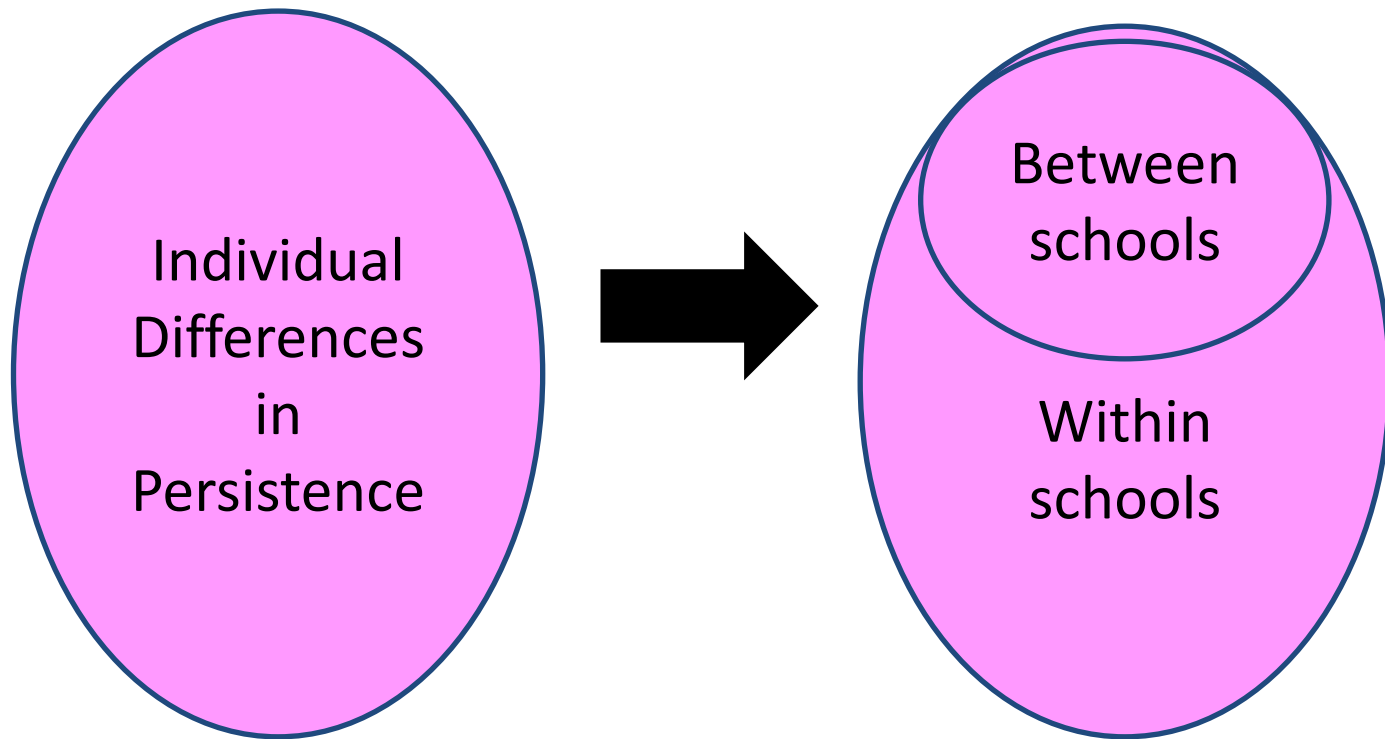
Cluster Position		1	2	3	4	5	6	7	8	9	10	11	12	13
	0	M1	R1	S1	R3	R4	R5	R6	R2	M2	S2	M3	R7	S3
	1	R1	S1	R3	R4	M2	R6	M3	M1	S2	R5	R7	S3	R2
	2	R3	R4	M2	S2	R5	R7	S3	S1	R6	M3	R2	M1	R1
	3	M3	R7	S3	R2	M1	R3	R4	R6	R1	S1	M2	S2	R5

R: Reading, M: Mathematics, S: Science

# Multilevel Decomposition

- In hierarchical sample structures (e.g. students within schools), the variation of measured variables can be decomposed in variance within and between groups.
- When individual differences in item position effects are modeled for hierarchical samples, these effects can also be decomposed by introducing the group level into the model.

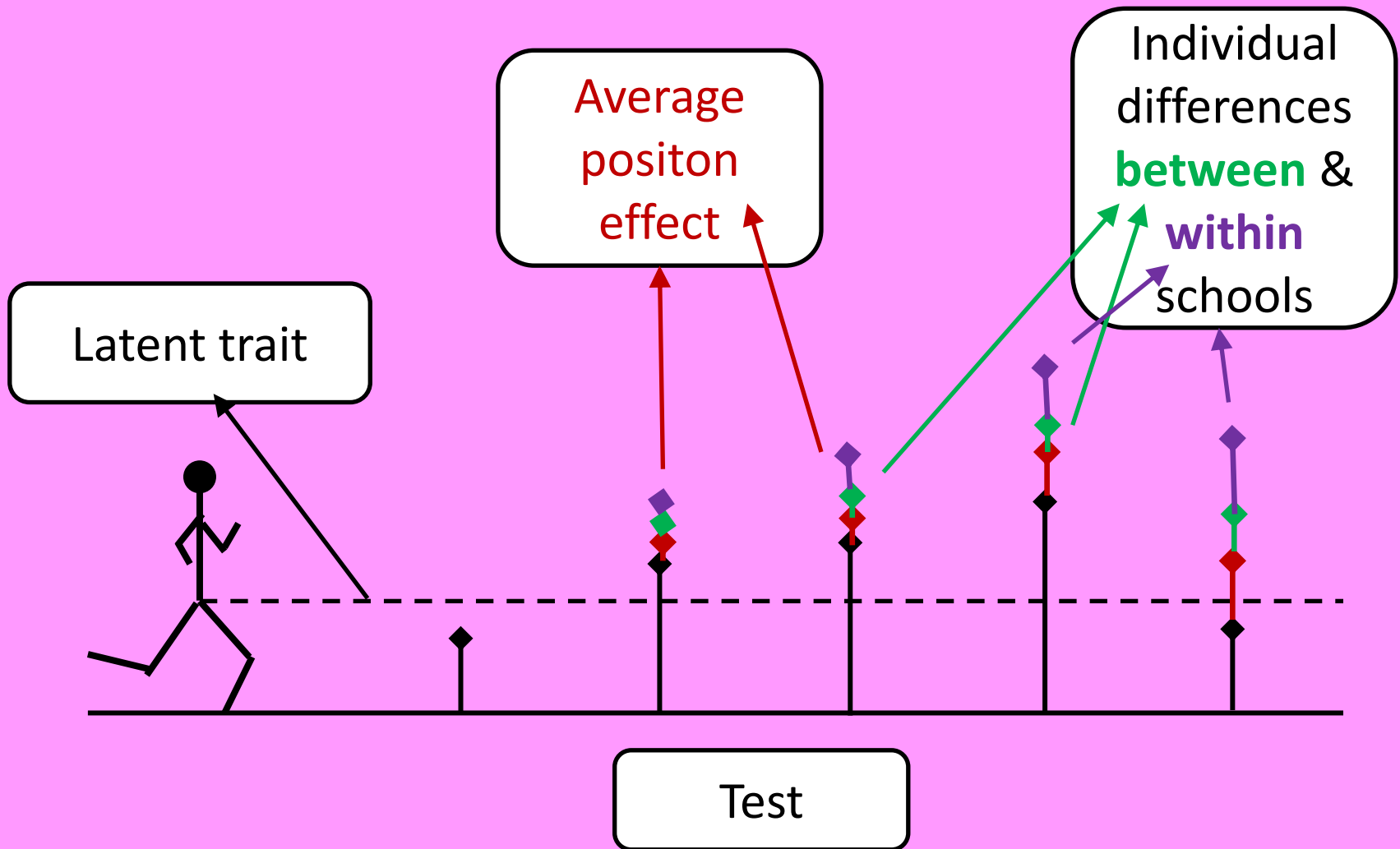
# Multilevel Decomposition



# Multilevel Decomposition

- Decomposing item position effects in educational assessments within and between schools might provide insights in the nature of these effects as well as in the structure of the school and student population.
- Proportions of variance between schools can be estimated for the ability measured as well as for item position effects.

# Multilevel Decomposition





# Research Questions

- Question 1: Is there a general item position effect, how large is it, and is it consistent across countries with different national performance levels?
- Question 2: Is there individual variance in item position effects (i.e. variation in persistence) in different countries?
- Question 3: How large is the proportion of variance (ICC) in persistence between schools?

# Research Questions

- Question 4: How are students' performance levels and persistence related within schools in different countries?
- Question 5: How are performance levels and persistence related on school level in different countries?
- Question 6: How are country-level results related to the countries' performance level?

# **B. PISA 2009 Reading assessment**

## **1. Introduction**

- PISA
- Multilevel decomposition
- Research questions

## **2. Model**

## **3. Results**

## **4. Discussion**

# Model

## Multilevel Model:

$$\text{logit}(Y_{ips}) = \theta_p^W + \theta_s^B - \beta_i + k_{pi}(\bar{\gamma} + \gamma_p^W + \gamma_s^B)$$

Person  $p$  ( $1, \dots, P$ )

School  $s$  ( $1, \dots, S$ )

Item  $i$  ( $1, \dots, I$ )

$\theta_p^W$  = ability level of student  $p$  (within-school)

$\theta_s^B$  = ability level of school  $s$  (between schools)

$\beta_i$  = difficulty of item  $i$

$k_{pi}$  = cluster position  $k$  ( $0, \dots, 3$ ) at which item  $i$  is presented to person  $p$

$\bar{\gamma}$  = average position effect across students and schools

$\gamma_p^W$  = position effect for student  $p$  (within school)

$\gamma_s^B$  = position effect for school  $s$  (between schools)

# Model

Level 1 (Response)

$$\text{logit}(Y_{ips}) = \theta_{ps} - \beta_i + k_{pi} \times \gamma_{ps}$$

Level 2 (Student)

$$\begin{aligned} \theta_{ps} &= \theta_p^W + \theta_s^B \\ \gamma_{ps} &= \gamma_p^W + \gamma_s^B \end{aligned} \quad \begin{pmatrix} \theta_p^W \\ \gamma_p^W \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\theta_p^W}^2 & \sigma_{\theta_p^W, \gamma_p^W} \\ \sigma_{\theta_p^W, \gamma_p^W} & \sigma_{\gamma_p^W}^2 \end{pmatrix} \right)$$

Level 3 (School)

$$\begin{pmatrix} \theta_s^B \\ \gamma_s^B \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ \bar{\gamma} \end{pmatrix}, \begin{pmatrix} \sigma_{\theta_s^B}^2 & \sigma_{\theta_s^B, \gamma_s^B} \\ \sigma_{\theta_s^B, \gamma_s^B} & \sigma_{\gamma_s^B}^2 \end{pmatrix} \right)$$

# Estimation

- All models were estimated with HLM,
- Student weights provided in the database were used as level 2 (student level) weighting variable.

# **B. PISA 2009 Reading assessment**

## **1. Introduction**

- PISA
- Multilevel decomposition
- Research questions

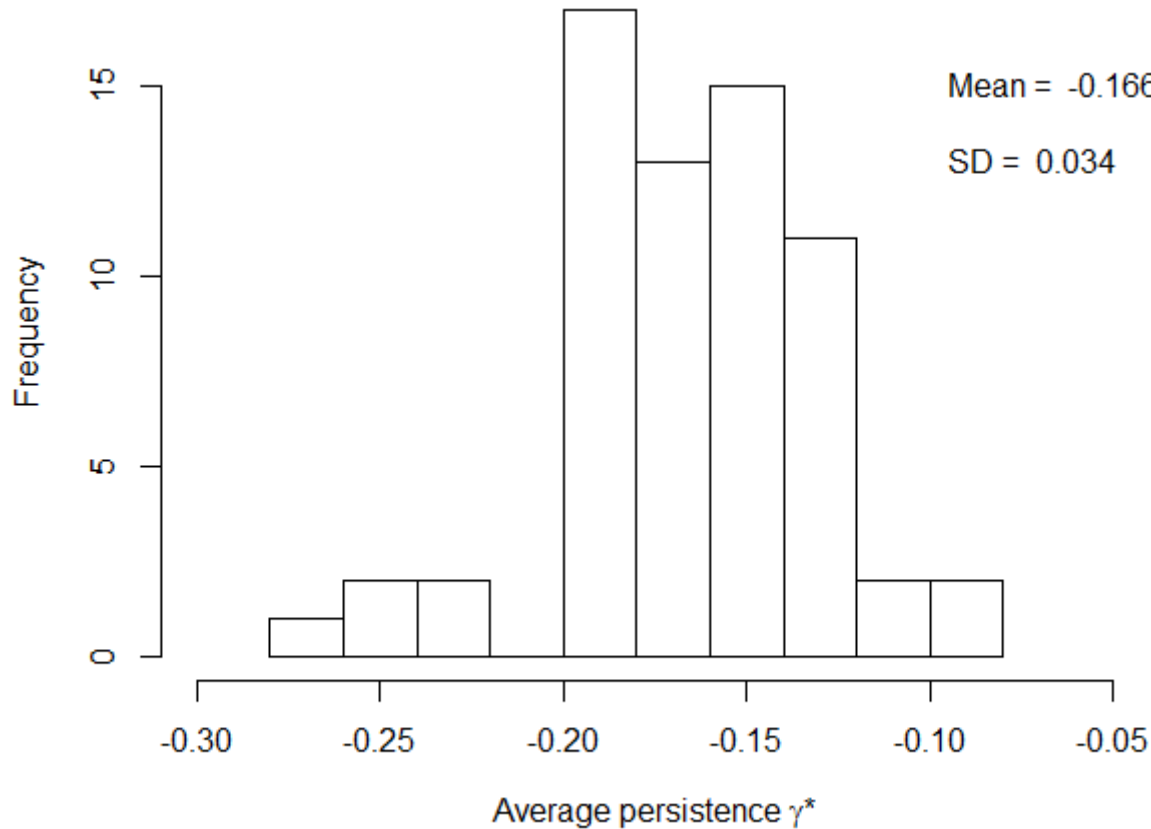
## **2. Model**

## **3. Results**

## **4. Discussion**

# Main Position Effects

## Fixed Effects of Item Cluster Position in 65 Countries



There is a negative effect of the cluster position in all countries, but the size varies substantially.

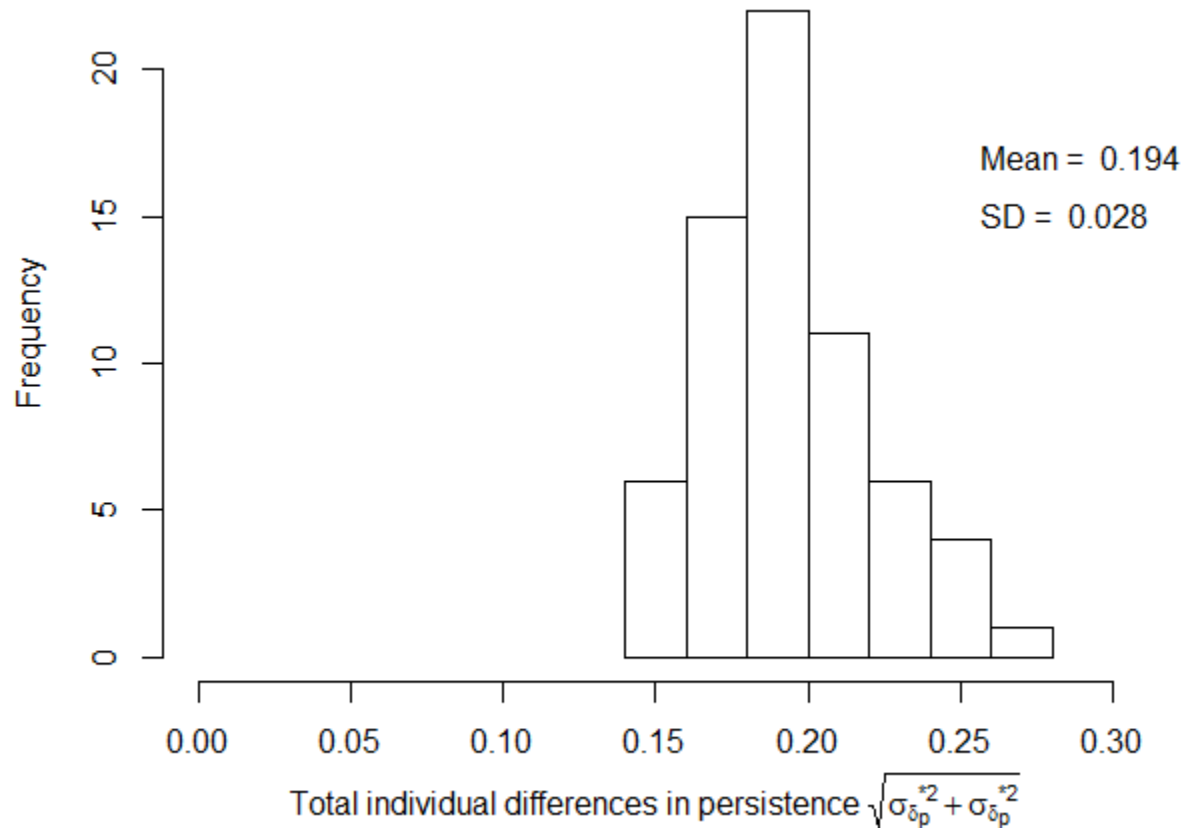


# Main Position Effects

Standardized persistence $\gamma^*$	Change in cluster positions	
	+ 1 cluster positions	+ 3 cluster positions
-0.277 ( <i>Greece</i> )	-.069	-.196
-0.167 ( <i>Average</i> )	-.042	-.122
-0.093 ( <i>Finland</i> )	-.023	-.069

# Random Position Effects

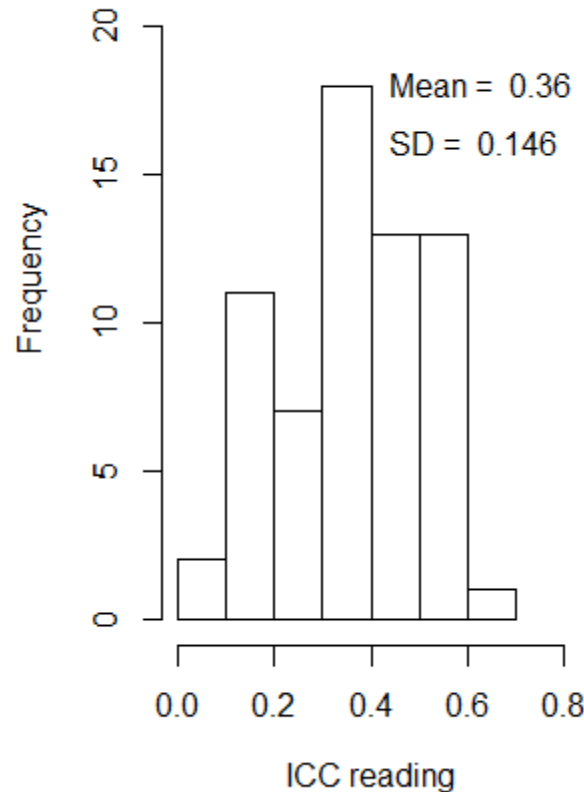
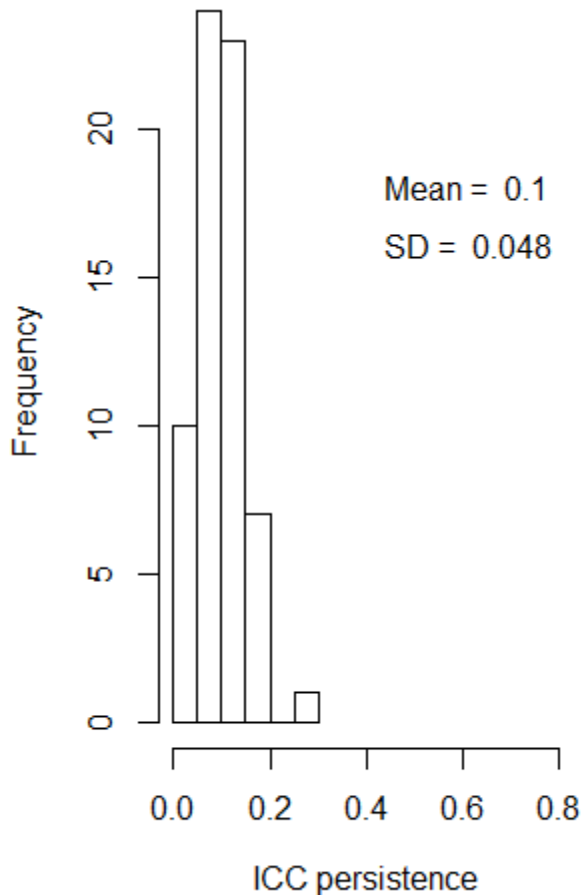
Overall individual differences in persistence



There are significant individual differences in persistence.

# Random Position Effects

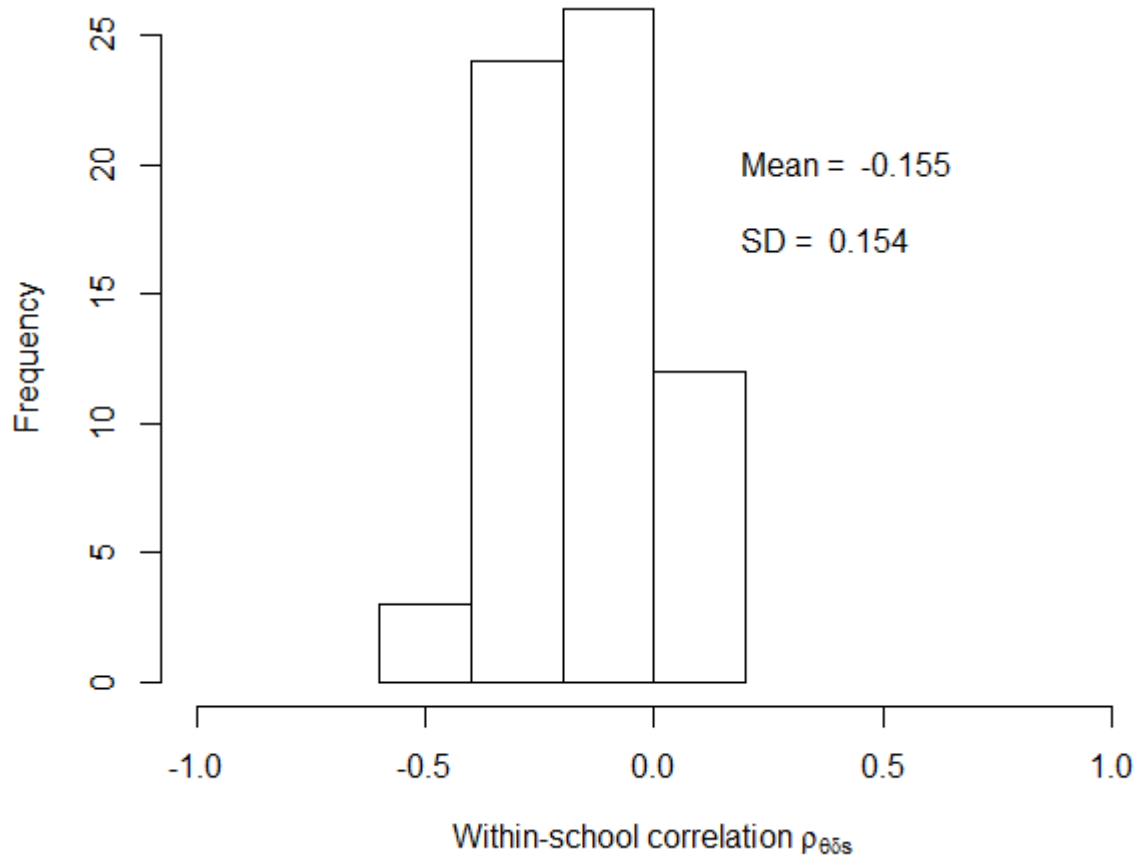
- ICC in reading vs. ICC in Persistence



A moderate amount of variance in persistence is located at the school level (less pronounced than for reading ability).

# Correlation Between Ability Level and Persistence

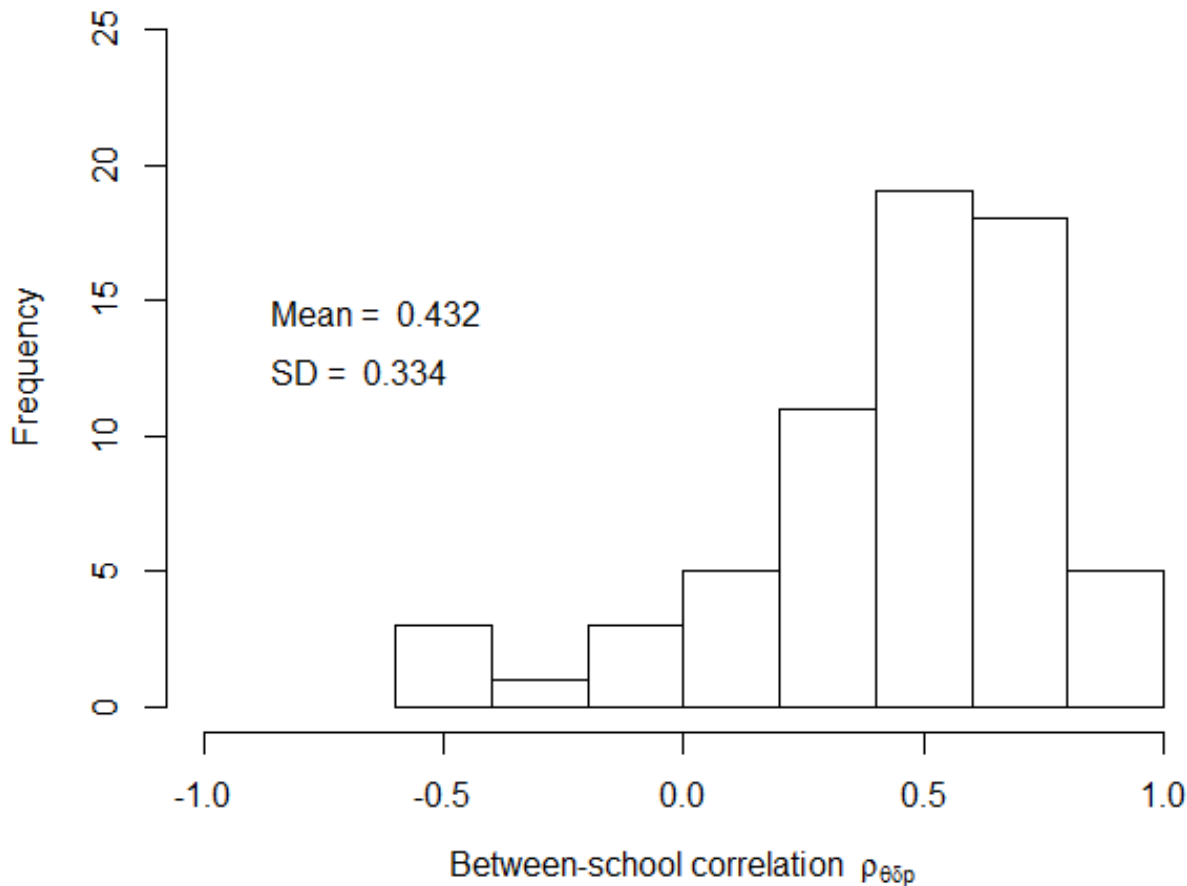
- Correlations within schools



Within schools, ability and persistence are correlated (slightly) negatively in most countries.

# Correlation Between Ability Level and Persistence

- Correlations between schools



Ability level and persistence level on school level are correlated positively in most countries.

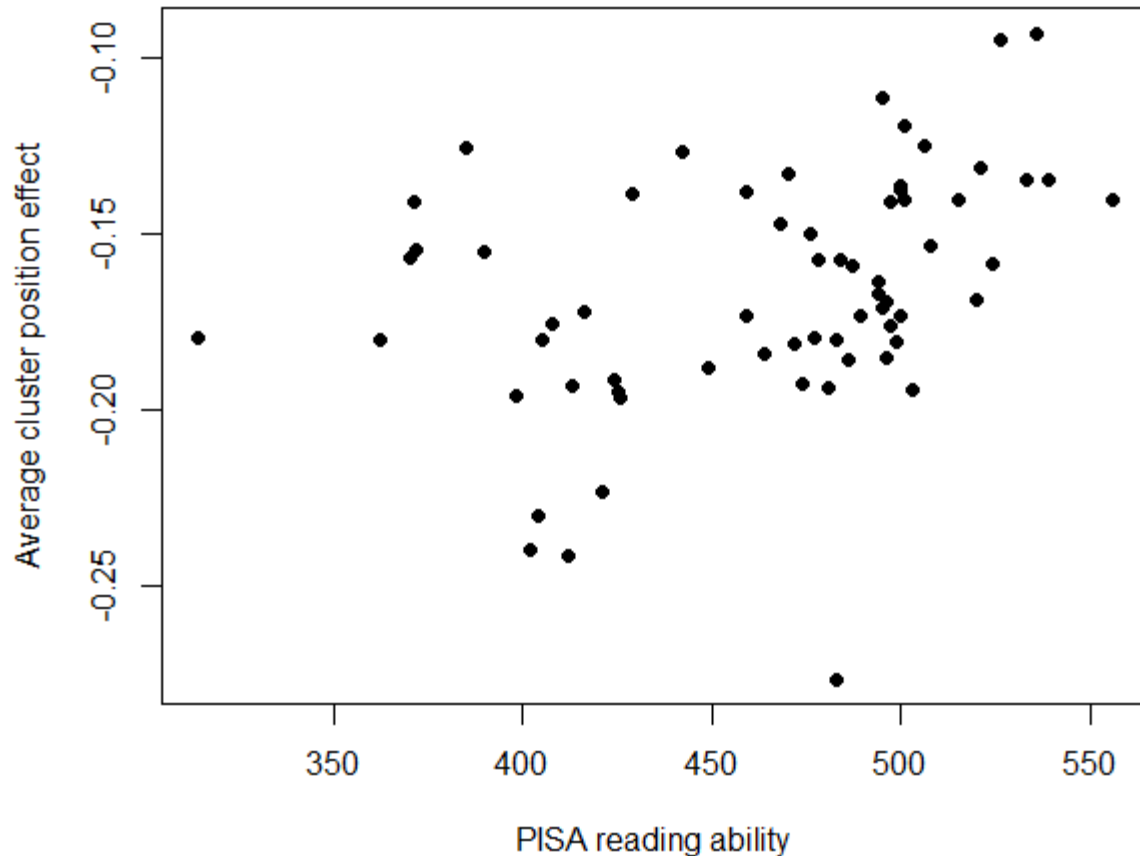
# Country Level Correlations with Performance Level

- Country level correlations between national reading scores and the model based results (N=65)

	Reading Score
Cluster position effect $\bar{\gamma}$	.36
Variance in persistence within schools $\sigma_{\gamma_p^W}^2$	-.54
Variance in persistence between schools $\sigma_{\gamma_p^B}^2$	-.06
Within-school correlation $r(\theta_p^W, \gamma_p^W)$	.56
Between-school correlation $r(\theta_s^B, \gamma_s^B)$	.55

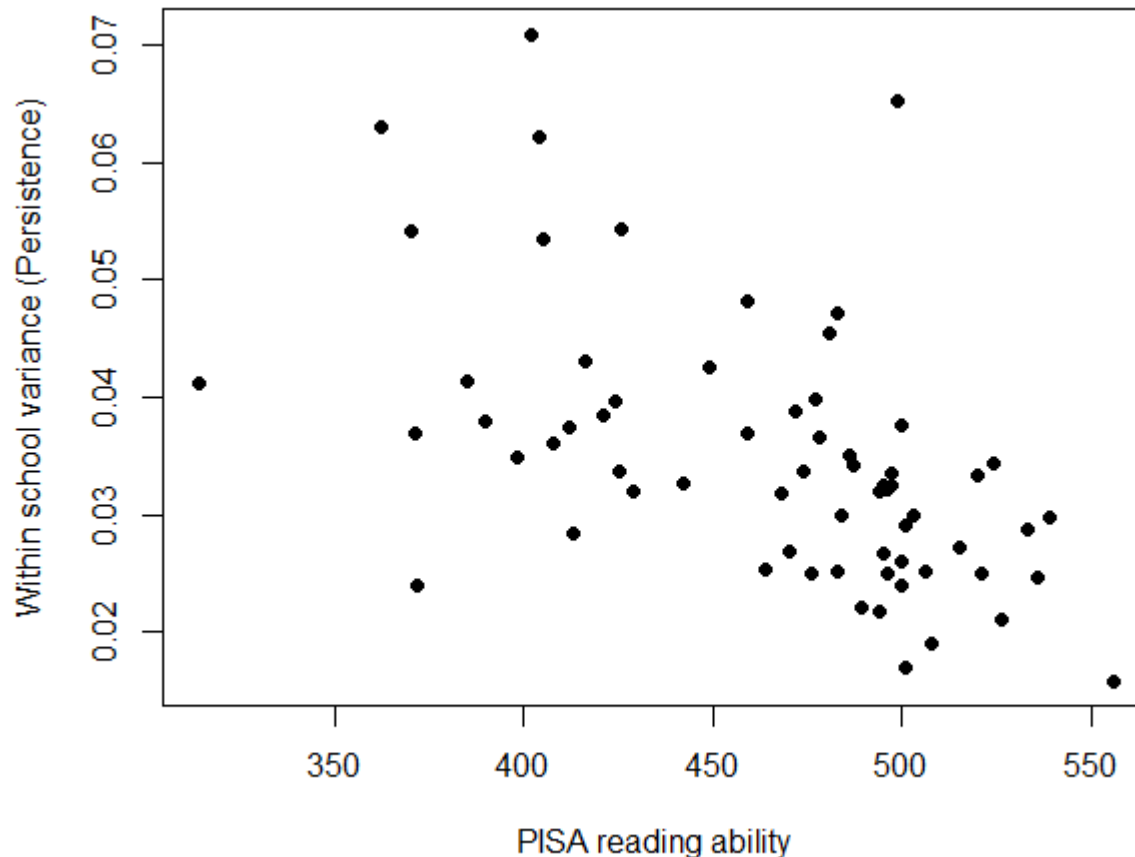
# Position Effects and National Performance Level

The strength of the position effect is related to the national performance level ( $r = .36$ )



# Position Effects and National Performance Level

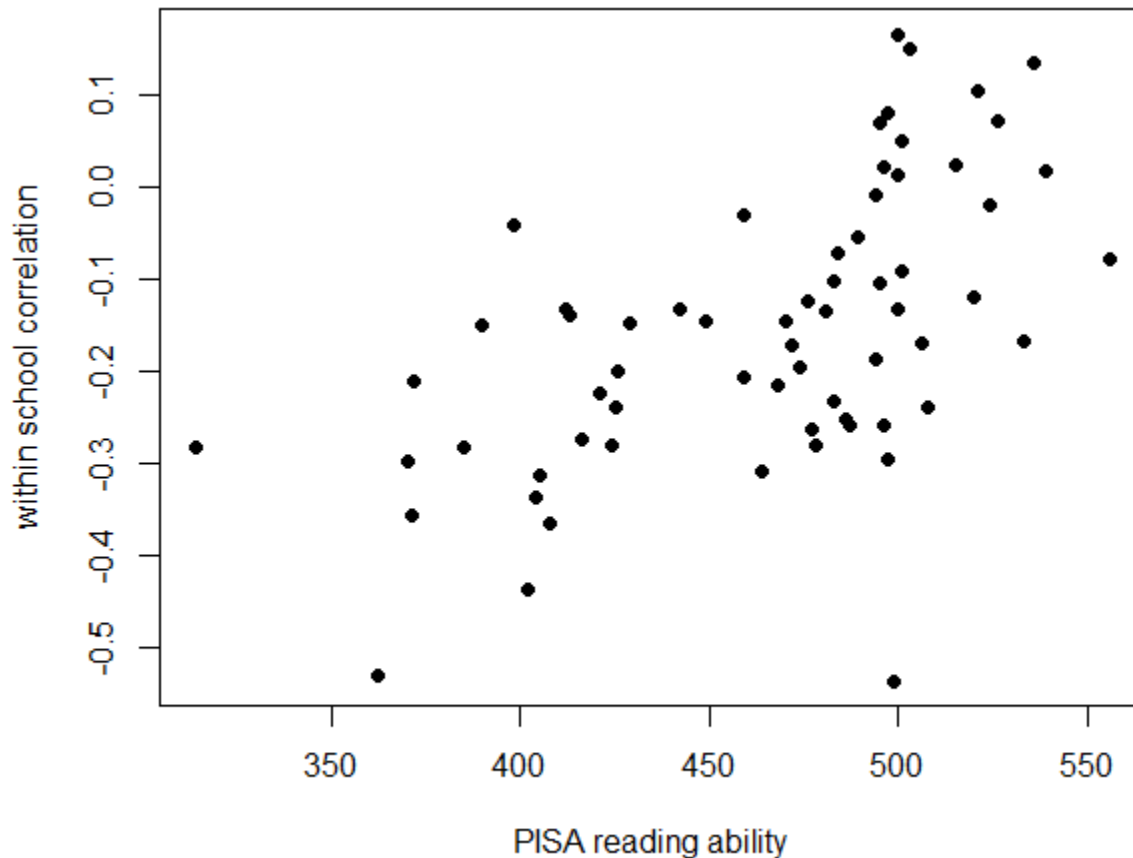
The variance of persistence on student level is negatively related to the national performance level ( $r = -.54$ )





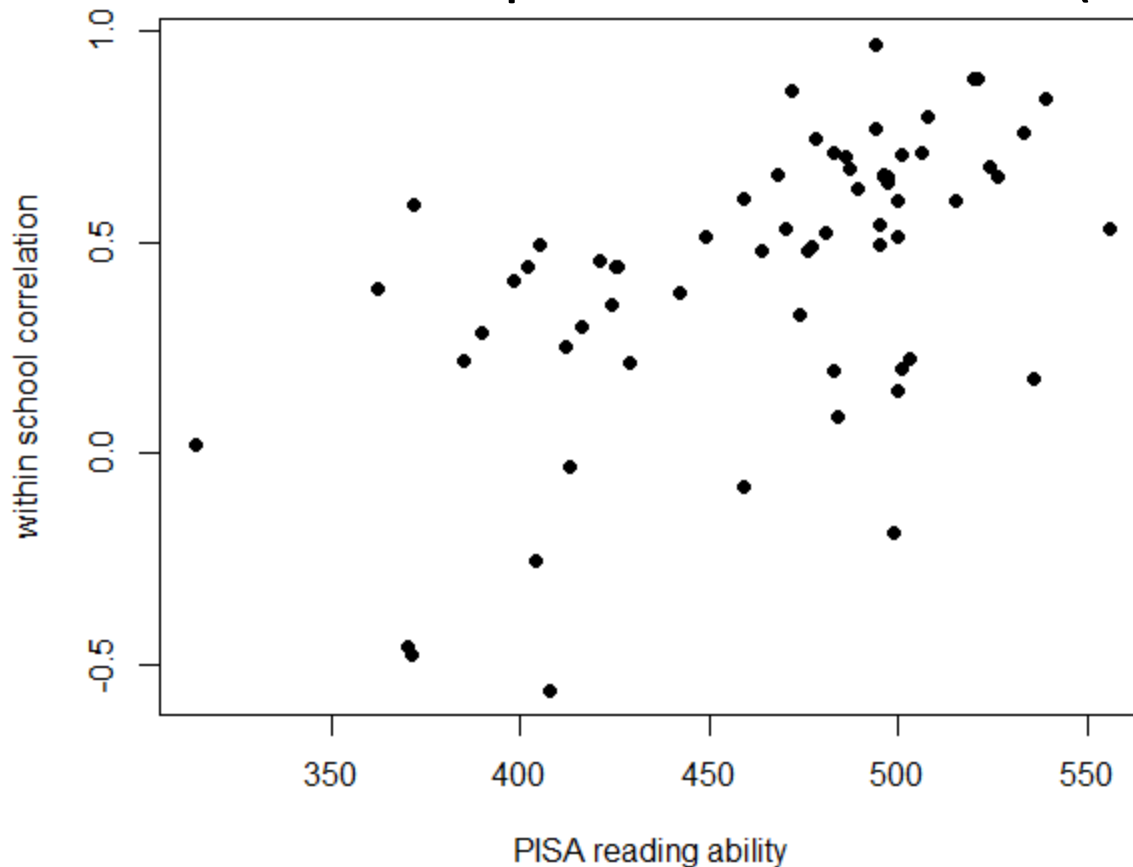
# Position Effects and National Performance Level

The correlation of ability and persistence within schools is related to the national performance level ( $r = .56$ )



# Position Effects and National Performance Level

The correlation of ability and persistence on school level is related to the national performance level ( $r = .55$ )



# Country Level Correlations with Performance Level

- The negative cluster position effect is less pronounced in higher performing countries.
- The variance in persistence between students and between schools is lower in higher performing countries.
- The correlation of ability and persistence within and between schools is higher in higher performing countries.

# **B. PISA 2009 Reading assessment**

## **1. Introduction**

- PISA
- Multilevel decomposition
- Research questions

## **2. Model**

## **3. Results**

## **4. Discussion**

# Discussion

- Results vary between countries, the only consistent finding is that the cluster position effect is negative.
- Include student level covariates (motivation)
- Include school level covariates
- We modeled a linear trend of cluster position, the appropriateness of this should be inspected on country level.

# Discussion

- Different response process variables (omission, no credit, full credit) are modeled in one variable. Maybe there are differences in the omitting process and accuracy process across students/schools/countries.
- Maybe there is an interaction between the item format (open ended/multiple choice) and the item position effect.
- Still far from understanding the mechanisms.

**your Thank attention for you!**

Position does matter!